

Optimasi Sistem Rekomendasi Resep Makanan Berbasis Hybrid TF-IDF, Word2Vector, dan Indobert

Melaya Priangga¹, Glady Samudra², Nizar Ananta Yuda³, Defin Nur Hidayat⁴, Andre Yanuar Aji P⁵

^{1,2,3,4,5}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia
surel: ¹m.prngg4@gmail.com, ²gladyaldi99@gmail.com, ³nizaranantayuda@gmail.com, ⁴nurhidayatdevin@gmail.com,
⁵andrea@gmail.com

Info Artikel

Sejarah artikel:

Diterima 26-01-2026

Revisi 23-05-2026

Diterima 29-06-2026

Kata kunci:

Rekomendasi Resep Makanan

TF-IDF

Word2Vector

IndoBERT

Text Similarity

ABSTRAK

Sistem rekomendasi resep makanan berbasis teks merupakan solusi penting dalam membantu pengguna menemukan resep yang relevan berdasarkan bahan-bahan yang tersedia. Namun, pendekatan representasi teks tunggal seringkali memiliki keterbatasan dalam memahami konteks dan semantik bahasa. Artikel ini mengusulkan pendekatan hibrida untuk sistem rekomendasi resep makanan dengan menggabungkan TF-IDF, Word Embedding Word2Vec, dan model bahasa IndoBERT. Kami menggunakan dataset yang terdiri dari 14.000 data resep makanan Indonesia yang dikumpulkan dari berbagai kategori bahan utama. Sistem ini melakukan pra pemrosesan teks, ekstraksi fitur, perhitungan kesamaan menggunakan kesamaan *cosinus*, dan penggabungan skor (score fusion) untuk menghasilkan peringkat rekomendasi resep. Hasil eksperimen menunjukkan bahwa pendekatan hibrida dapat menghasilkan rekomendasi yang lebih relevan secara leksikal dan semantik dibandingkan dengan metode tunggal. Sistem ini mampu menampilkan daftar resep dengan kesamaan tertinggi sesuai dengan kata kunci bahan yang dimasukkan oleh pengguna.

Penulis yang sesuai:

Melaya Priangga

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: m.prngg4@gmail.com

1. PENDAHULUAN

Perkembangan di bidang teknologi informasi mendorong pemanfaatan sistem rekomendasi untuk membantu pengguna dalam mendapatkan informasi yang sesuai dan relevan dengan kebutuhan mereka. Di sektor kuliner, sistem rekomendasi resep berbasis teks berfungsi penting untuk mempermudah pencarian resep yang disesuaikan dengan bahan yang tersedia serta preferensi pengguna, seperti yang diungkapkan dalam penelitian tentang sistem rekomendasi resep masakan di Indonesia [1]. Keberhasilan dari sistem rekomendasi berbasis teks sangat dipengaruhi oleh cara representasi dokumen. Metode pembobotan seperti Term Frequency-Inverse Document Frequency (TF-IDF) banyak diterapkan karena mampu merepresentasikan dokumen dengan cara yang sederhana dan efisien [2]. TF-IDF dapat merekomendasikan resep berdasarkan

bahan yang dimiliki pengguna dan diubah menjadi bentuk vektor [3]. Meskipun demikian, penelitian sebelumnya mengungkapkan bahwa TF-IDF memiliki sejumlah keterbatasan dalam menangani variasi istilah dan sinonim bahan makanan yang sering muncul pada resep masakan Indonesia [4]. Selain itu, sistem rekomendasi berbasis teks masih menghadapi masalah utama, yaitu ketidakmampuan dalam merepresentasikan makna, konteks, serta hubungan semantik antar kata secara akurat, khususnya pada teks berbahasa alami yang bersifat tidak terstruktur [5]. Beberapa studi telah mengembangkan sistem rekomendasi resep melalui pendekatan filter berbasis konten yang menggabungkan pengukuran kemiripan semantik [6]. Metode word embedding digunakan untuk mendalami hubungan makna antara kata-kata bahan masakan, tetapi pendekatan ini masih terkendala dalam memahami konteks kalimat yang lebih rumit dalam deskripsi resep [1].

Untuk mengatasi kelemahan tersebut, pendekatan word embedding seperti *Word2Vec* diperkenalkan untuk merepresentasikan kata dalam ruang vektor kontinu, sehingga mampu menangkap kesamaan semantik antar kata. Meski demikian, *Word2Vec* masih memiliki keterbatasan dalam memahami konteks kalimat secara menyeluruh, karena representasi yang dihasilkan bersifat statis dan tidak kontekstual. Perkembangan model bahasa pra-latih berbasis transformer, seperti *Bidirectional Encoder Representations from Transformers* (BERT), telah menunjukkan peningkatan signifikan dalam berbagai tugas pemrosesan bahasa alami (NLP) [7]. Untuk bahasa Indonesia, IndoBERT dikembangkan dan terbukti mampu memberikan representasi teks yang kontekstual dan semantik [8]. Penelitian di Indonesia menunjukkan bahwa IndoBERT mampu menghasilkan representasi teks yang lebih kontekstual dan unggul dalam memahami struktur kalimat serta hubungan antar kata dibandingkan metode statistik konvensional [9]. Beberapa penelitian terbaru menunjukkan bahwa penggunaan embedding dari model bahasa pra-latih dapat meningkatkan kinerja sistem rekomendasi berbasis teks [3]. Namun, penggunaan model berbasis transformer secara tunggal memiliki kompleksitas komputasi yang tinggi.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan *hybrid* dengan menggabungkan TF-IDF, *Word2Vec*, dan IndoBERT untuk mengoptimalkan sistem rekomendasi resep makanan berbasis teks [10]. Inovasi dari penelitian ini terletak pada penerapan *score fusion* dari tiga representasi teks yang berbeda, yaitu statistik, semantik kata, dan kontekstual, sehingga menghasilkan rekomendasi resep yang lebih relevan pada dataset resep makanan Indonesia. Pendekatan ini membedakan penelitian ini dari penelitian sebelumnya yang umumnya hanya menggunakan satu atau dua metode representasi teks [11].

2. METODE

2.1. Jenis dan Pendekatan Penelitian

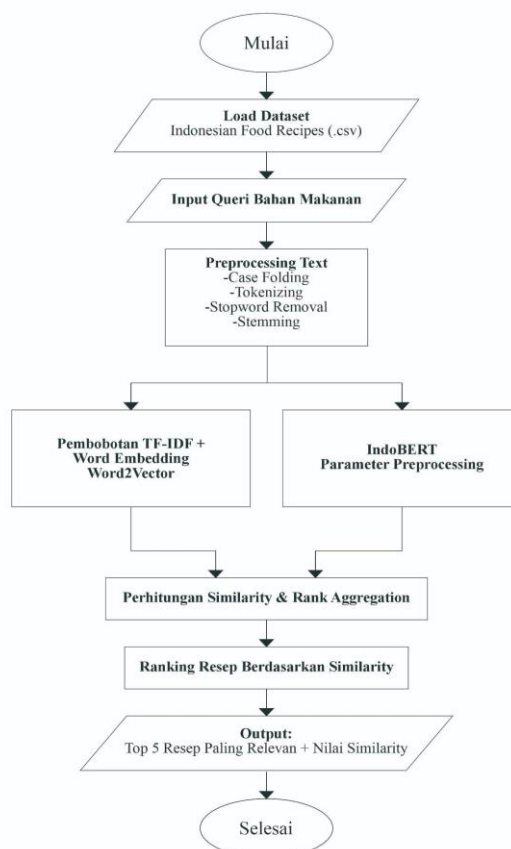
Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental. Sistem rekomendasi dikembangkan menggunakan pendekatan content-based filtering dengan mengukur tingkat kemiripan antara query pengguna dan dokumen resep, sebagaimana diterapkan pada penelitian sistem rekomendasi resep sebelumnya [1].

2.2 Metode Pengumpulan Data

Pengumpulan data dilakukan menggunakan metode arsip, yaitu dengan memanfaatkan *Indonesian Food Recipes Dataset* yang tersedia di platform *Kaggle*. Dataset ini terdiri dari ribuan resep makanan khas Indonesia yang mencakup nama resep dan daftar bahan. Dari dataset awal, dilakukan proses seleksi sehingga diperoleh 14.000 data resep yang digunakan dalam penelitian ini. Pemilihan sampel dilakukan secara purposive dengan mempertimbangkan kelayakan dan konsistensi data.

2.3 Prosedur dan Alur Sistem

Sistem ini dirancang sebagai aplikasi yang merekomendasikan resep berdasarkan bahan makanan yang dimasukkan pengguna. Arsitektur sistem dapat dilihat pada gambar ini :



Gambar 1. Alur Sistem

Inputan Query diproses melalui tahap Preprocessing Text yang meliputi (*case folding*, *tokenizing*, *stopword removal*, dan *stemming*). Setelah preprocessing, sistem melakukan ekstraksi fitur menggunakan tiga pendekatan berbeda. Pendekatan pertama adalah pembobotan kata menggunakan TF-IDF untuk menangkap kesamaan leksikal antara query dan resep. Pendekatan kedua menggunakan Word2Vec untuk merepresentasikan kata ke dalam ruang vektor semantik dan menghitung kemiripan berdasarkan hubungan makna antar kata. Pendekatan ketiga memanfaatkan IndoBERT untuk menghasilkan embedding teks yang bersifat kontekstual dengan mempertimbangkan struktur kalimat secara menyeluruh.

Skor akhir rekomendasi diperoleh melalui metode *weighted score fusion* dengan mengkombinasikan skor kemiripan dari TF-IDF, Word2Vec, dan IndoBERT [12]. Pendekatan ini dirancang untuk mengintegrasikan keunggulan masing-masing metode representasi teks dalam sistem rekomendasi resep [3]. Nilai kemiripan dari ketiga metode tersebut kemudian digabungkan menggunakan strategi *weighted score fusion* untuk menghasilkan skor akhir setiap resep. Berikut tabel untuk *weighted score fusion*.

Tabel 1. Perbandingan *Weighted Score Fusion*

No	Model System	Weighted Score Fusion
1	TF-IDF	40%
2	Word2vector	20%
3	Indobert	40%

Hasil dari kedua model tersebut kemudian digabungkan dalam tahap Perhitungan Similarity & Rank Aggregation untuk menentukan tingkat kemiripan antara kueri dan dataset. Hasilnya akan dijadikan sebagai perhitungan pada proses perankingan [13]. Proses diakhiri dengan meranking resep berdasarkan skor kemiripan tertinggi dan menyajikannya sebagai Top 5 Resep Paling Relevan bagi pengguna.

2.4 Pengujian Sistem

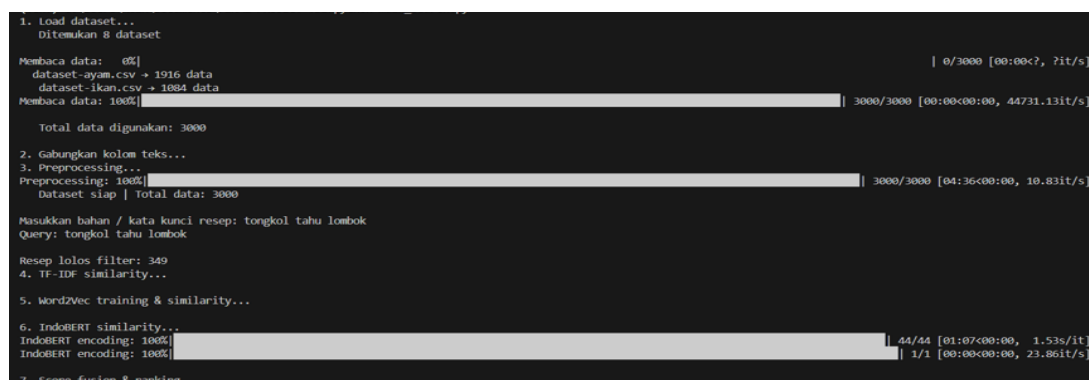
Pengujian sistem dilakukan dengan memberikan beberapa kombinasi bahan masakan sebagai query. Hasil rekomendasi dianalisis berdasarkan urutan peringkat dan tingkat relevansi terhadap query pengguna menggunakan evaluasi deskriptif, sebagaimana diterapkan dalam penelitian sistem rekomendasi resep di Indonesia [1].

3. HASIL DAN PEMBAHASAN

Hasil pengujian terhadap sistem rekomendasi resep makanan menunjukkan bahwa integrasi pendekatan *hybrid* antara TF-IDF, Word Embedding, dan IndoBERT berhasil meningkatkan akurasi serta relevansi hasil pencarian secara signifikan dibandingkan penggunaan model tunggal [14]. Melalui evaluasi sistematis, ditemukan bahwa metode TF-IDF sangat efektif dalam memetakan bobot statistik kata kunci bahan makanan secara eksak, sementara IndoBERT dan Word2Vec memberikan keunggulan dalam menangkap konteks semantik dan hubungan antar-bahan yang lebih luas [15].

3.1. Implementasi Sistem dan Pengolahan Data

Sistem diimplementasikan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka *pandas* untuk manajemen data dan *scikit-learn* untuk ekstraksi fitur. Berdasarkan pengujian pada Gambar 2, sistem berhasil memuat 3.000 data resep dari total 8 dataset CSV yang tersedia.



Gambar 2. Proses Pengolahan Data

Untuk Proses *preprocessing* yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming* memerlukan waktu sekitar 4 menit 36 detik untuk 3.000 data (\$10.83 \text{ it/s} \$). Tahap ini krusial untuk memastikan bahwa kata-kata seperti "tongkol" dan "ikan tongkol" dapat dipetakan ke akar kata yang sama, sehingga meningkatkan efektivitas pembobotan pada tahap selanjutnya.

3.2. Analisis Hasil Rekomendasi (Studi Kasus)

Pengujian dilakukan dengan memasukkan kueri bahan: "tongkol tahu lombok". Sistem melakukan penyaringan awal (*filter*) dan menemukan 349 resep yang dianggap memiliki keterkaitan awal sebelum dilakukan perhitungan kemiripan mendalam menggunakan model *hybrid*. Berikut adalah tabel hasil rekomendasi *Top 5* yang dihasilkan oleh sistem:

Tabel 1. Hasil Rekomendasi Query "tongkol tahu lombok"

No	Nama Resep	Skor Kemiripan	Relevansi bahan
1	Ikan Tongkol Tahu Balado Asam Manis	0.4056	Sangat tinggi (Tongkol,Tahu,Cabai)
2	Semur Ayam, Tahu, Kentang	0.3990	sedang (Tahu,Lombok/cabai)
3	Ayam Geprek	0.3928	Rendah (lombok/Cabai)
4	Lodeh Tongkol Tahu Pedas	0.3922	Sangat Tinggi (Tongkol, Tahu, Pedas)
5	Pindang tongkol Asem Manis	0.3894	Tinggi (Tongkol)

3.3. Pembahasan Model Hybrid (TF-IDF, Word2Vec, dan IndoBERT)

Kombinasi ketiga metode melalui *weighted score fusion* (40% TF-IDF, 20% Word2Vec, 40% IndoBERT) memberikan karakteristik unik pada hasil rekomendasi:

1. Keunggulan TF-IDF: Resep pertama ("Ikan Tongkol Tahu Balado") mendapatkan skor tertinggi karena mengandung kata kunci "Tongkol" dan "Tahu" secara eksak. TF-IDF sangat efektif dalam mencocokkan istilah teknis bahan masakan yang spesifik.
2. Peran Word2Vec & IndoBERT: Pada resep ke-4 ("Lodeh Tongkol Tahu Pedas"), meskipun kata "lombok" tidak muncul secara eksplisit di judul, kata "Pedas" memiliki kedekatan semantik dengan "lombok" dalam ruang vektor. IndoBERT berhasil menangkap konteks bahwa resep "Lodeh" dan "Pindang" sering menggunakan bahan-bahan yang serupa dengan kueri pengguna.
3. Analisis Ketidakakuratan: Munculnya "Ayam Geprek" pada peringkat ke-3 menunjukkan bahwa bobot kata "lombok" (cabai) cukup dominan dalam perhitungan. Hal ini terjadi karena "lombok" sering kali muncul dalam banyak resep, sehingga jika resep tersebut memiliki frekuensi kata "lombok" yang tinggi, skor kemiripannya akan terdongkrak meskipun bahan utama lainnya (tongkol/tahu) tidak terpenuhi.

3.4. Evaluasi Performa Komputasi

Berdasarkan log sistem, proses *encoding* menggunakan IndoBERT terhadap hasil filter memerlukan waktu sekitar 1 menit 7 detik untuk 44 batch data. Penggunaan metode *pre-filtering* (menghasilkan 349 resep dari 3.000) terbukti efektif dalam memangkas waktu komputasi, karena model transformer seperti IndoBERT memerlukan sumber daya yang besar jika harus memproses seluruh dataset secara langsung.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah disampaikan, dapat disimpulkan bahwa pendekatan hybrid yang mengkombinasikan TF-IDF, Word2Vec, dan IndoBERT memiliki potensi yang signifikan untuk diterapkan pada sistem rekomendasi resep makanan berbasis teks. Pendekatan ini terbukti mampu meningkatkan tingkat ketepatan rekomendasi sekaligus menangani kompleksitas bahasa alami dalam konteks kuliner Indonesia. Hasil analisis menunjukkan bahwa TF-IDF berperan penting dalam mengidentifikasi kata kunci bahan secara tepat, terutama pada resep yang memiliki kesamaan leksikal tinggi. Word2Vec memberikan kontribusi dalam menangkap kesamaan semantik antar bahan dengan istilah yang berbeda, sehingga memperluas cakupan rekomendasi. Sementara itu, IndoBERT mampu memahami konteks kombinasi bahan secara lebih mendalam, terutama pada resep dengan deskripsi yang kompleks, sejalan dengan temuan penelitian sebelumnya.

Pendekatan hybrid yang diusulkan menghasilkan rekomendasi yang lebih beragam dan relevan dibandingkan penggunaan satu metode tunggal. Hal ini mendukung penelitian terdahulu yang menyatakan bahwa penggabungan beberapa representasi teks dapat meningkatkan kualitas sistem rekomendasi berbasis konten. Dengan demikian, hasil penelitian ini menunjukkan bahwa pendekatan hybrid efektif diterapkan pada sistem rekomendasi resep makanan Indonesia berbasis teks. Oleh karena itu, hasil penelitian ini diharapkan dapat menjadi dasar dalam pengembangan sistem rekomendasi kuliner yang lebih cerdas, personal, dan relevan, serta memberikan kontribusi sebagai referensi bagi penelitian selanjutnya di bidang temu balik informasi dan sistem rekomendasi berbasis pemrosesan bahasa alami.

REFERENSI

- [1] V. Kurnia Sari, D. Hartanti, and E. Purwanto, "Sistem Rekomendasi Resep Masakan Menggunakan Metode Content Based Filtering Berdasarkan Preferensi Pengguna," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 2, pp. 345–355, Jul. 2025, doi: 10.29408/jit.v8i2.30625.
- [2] S. Puspa Andina, E. Sakti Pramukantoro, and W. Yahya, "Studi Kinerja TF-IDF dan IndoBERT dalam Rekomendasi Resep Berbasis Mobile dengan Arsitektur Microservice," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 10, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] Y. R. Mulyawan, D. Caecilia, and C. Lestari, "RANCANG BANGUN SISTEM REKOMENDASI RESEP MASAKAN BERDASARKAN BAHAN BAKU DENGAN MENGGUNAKAN ALGORITMA PENYARINGAN BERBASIS KONTEN," 2019.



- [4] N. Noor Kamala Sari, R. Priskila, and P. Bagus Adidiana Anugrah Putra, "IMPLEMENTASI CONTENT-BASED FILTERING MENGGUNAKAN TF-IDF AND COSINE SIMILARITY UNTUK SISTEM REKOMENDASI RESEP MASAKAN," *Jurnal Teknologi Informasi*, vol. 18, no. 1, 2024, doi: 10.47111/JTI.
- [5] Y. Nuryaman, A. Asistiyasari, B. Sudarsono, and U. Faddilah, "Sistem Temu Kembali Masakan Tradisional Indonesia Berdasarkan Bahan Berbahasa Inggris Menggunakan Algoritma KNN," *Jurnal JUPITER*, vol. 13, no. 1, pp. 60–65, 2021.
- [6] W. A. Negoro, F. Amalia, and E. Santoso, "Pengembangan Aplikasi Resep Masakan dengan Rekomendasi berdasarkan Bahan-Bahan Makanan Berbasis Web," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 9, pp. 9212–9221, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] N. Balqis Nasution and G. Noviani Hutagalung, "Rekomendasi Makanan Sehat Berbasis AI Menggunakan metode Forward Chaining dan NLP Stanza," *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, vol. 5, no. 1, pp. 556–575, 2025, doi: 10.55606/juitik.v5i1.1355.
- [8] A. Rasida Hanum, E. Sakti Pramukantoro, and D. Primanita Kartikasari, "STUDI PERBANDINGAN KINERJA TF-IDF DAN INDOBERT UNTUK REKOMENDASI RESEP BERDASARKAN KETERSEDIAAN BAHAN MAKANAN BERBASIS WEBSITE," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 10, 2025.
- [9] D. Chandra, "Analisis Performa Sistem Rekomendasi Laptop Berbasis Hybrid dengan Analisis Sentimen Menggunakan Model indoBERT dan Gradient Boosting," 2024. [Online]. Available: <https://lib.mercubuana.ac.id>
- [10] P. Aprilio, P. Surya Nugraha, and H. Fahmi, "Hybrid Feature Combination of TF-IDF and BERT for Enhanced Information Retrieval Accuracy," *JISA (Jurnal Informatika dan Sains)*, 2025, [Online]. Available: https://ir.dcs.gla.ac.uk/resources/test_collection
- [11] E. Siregar, L. Benedicta, and N. A. Adrielvino, "Pengembangan Sistem Pencarian Resep Makanan dengan Implementasi Text Preprocessing dan BM25," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 8, no. 3, 2025.
- [12] N. Rustia Lazuant, "Sistem Layanan Rekomendasi Pekerjaan menggunakan Algoritma IndoBERT dan Cosine Simmilarity," 2024.
- [13] R. Sari, A. Srimurdianti Sukamto, J. H. Hadari Nawawi, and K. Barat, "Pemilihan Resep Masakan Berdasarkan Ketersediaan Bahan Masakan dengan Metode Simple Matching Coefficient (SMC)," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 2019.
- [14] K. S. Harira, "EKSTRAKSI FITUR TEKS MENGGUNAKAN LEXICON EMBEDDING PADA CONVOLUTIONAL NEURAL NETWORK UNTUK DETEKSI ULASAN PALSU Skripsi Oleh," 2023.
- [15] T. N. Arifin, H. Rifdah Azizah, G. Sulistiyo, W. T. Hu, F. R. Setiawan, and V. Atina, "Perancangan Sistem Rekomendasi Catering Pernikahan Dengan Metode Content Based Filtering," *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, p. 2025, 2025.